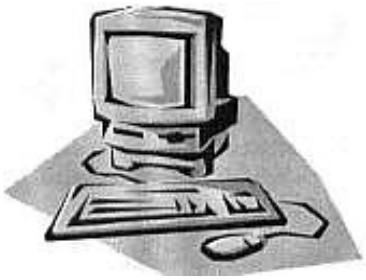
COMPUTER CORNER



What *is* The Cloud and Why Should I Care?

Strategy for Genealogists — Part 1

Guest Column by Dick Eastman

*Dick Eastman has been involved in genealogy for more than 30 years. He has worked in the computer industry for more than 40 years in hardware, software, and managerial positions. Dick is well known to many AGS members, both from his on-line genealogy blog (*Eastman's Online Genealogy Newsletter) *as well as his presentation of an all-day genealogy seminar in Albuquerque in April of last year.*

This April in the Plus Edition of his on-line newsletter, Mr. Eastman explored the concept of cloud computing for genealogists in a 3-part series. Computer Corner has touched on this subject in prior columns; however, Eastman has covered it in much more detail. Viewing the full original series may take a subscription to Plus on www.eogn.com. *We have received permission from the author to re-print the series, which we have edited into two parts, the first of which appears below.*

Mike Blackledge

The newest technology in computers these days is called "cloud computing. " However, the buzzword is so new that many experienced computer users do not yet understand the term. In fact, "the cloud" can be different things to different people. This article [first part] explains what the cloud is, and the second part will describe using the cloud for genealogy purposes.

In recent weeks, Ancestry.com had outages with its RootsWeb/USGenWeb/WorldConnect web pages. Some users declared, "So much for the cloud and trusting others with your data." However, that statement shows a lack of familiarity with cloud computing. While Ancestry uses industrial grade computers as servers to deliver information on the World Wide Web, they obviously are not using cloud computing methods for affected services that suffered weeks of outage.

NOTE: Most cloud-based services provide 99.999% or better uptime. An article last year in *Network World* states that "Amazon's Elastic Compute Cloud (EC2) recorded 2.41 hours of downtime across 20 outages in 2014, meaning it was up and running 99.9974% of the time." The same article also states, "Perhaps even more eye-catching is the uptime of Google Cloud Platform's storage service, which experienced 14 minutes of downtime in all of 2014, according to CloudHarmony. That's good for a 99.9996 uptime percentage."

Such statistics are common for true cloud-based services running in multiple data centers. It is the intent of this article to reduce some of the confusion about the term "cloud computing."

Computers are marvelous things. We can sit at home with a desktop computer or use a laptop when traveling or even use a handheld computer while sitting in a city park. We can watch YouTube videos, read and write email, check the latest news and weather reports, research a family tree, plot the best travel route to Poughkeepsie, or perform any of dozens of other tasks, regardless of our location. We can do this because we are connected to a monstrous collection of computers and computer accessories that include routers, switches, hubs, and miles and miles of cabling.

In the traditional world of desktop applications, data is usually stored on a computer's hard drive. In "the old days" of a few years ago, I could go on vacation and leave my computer at home but then could not access my email, photos, or any of my data. In the new world of cloud services, my email and all my data are safely and securely stored online, that is, in the cloud. I can get to it by using a web browser from any computer that's connected to the Internet. Even better, that information is available to me essentially all the time.

In many cases, I can even get to it from a so-called "smartphone," that is, a cellphone with a built-in computer that usually runs on an Android or Apple operating system.

The hardware that lets us compute anyplace on any device is invisible to us, and most of us do not understand how it all works. However, our computers, even our tiny cell phone "smartphones," are actually plugged into the collective power of thousands of computers that serve all this information to us from far-away computer systems distributed around the world. It's almost like having a massive supercomputer at your beck and call, thanks to the Internet.

There may be hundreds of thousands of hardware boxes involved along with millions of services available on the Internet. This collection of hardware, software, and data is collectively called "the cloud."

In fact, some people think that any web server that is available on the World Wide Web is part of "the cloud." However, that impression is incorrect; a web server can be any computer, large or small, that serves up data to the web where people can see it. For instance, the web server that is displaying this [on-line newsletter] to you is a single web server installed in one data center. If that web server stops working at some point, the web pages you read will not be available to you.

Definitions are not rigidly defined. The term "cloud" can be taken to mean many different things. However, most data processing professionals will tell you that a single web server installed in a single data center does not qualify as a cloud-based service.

Generally speaking, a true cloud service uses a number of servers that are running in multiple data centers located in different locations around the world. Perhaps even more important, there are always two or more servers in operation with the same information loaded onto each server, and then the workload is distributed amongst these multiple servers in multiple locations. Companies that offer this distributed and redundant service include Amazon Web Services, RackSpace, Microsoft Cloud, AT&T Synaptic Compute, and a bunch of other companies. I suspect you will be surprised at how many companies are in the "rent-some-cloud-space" business. Most of these companies offer services to commercial and non-profit organizations, such as to FamilySearch, Ancestry, MyHeritage, the New England Historic Genealogical Society, and many others. In turn, those organizations use these huge cloud services to serve the needs of their own users.

Most of us think of these services as "the web," and that term certainly is correct. However, "the cloud" consists of both web sites that display information AND servers where we can make backups of our own files AND programs that we can run AND much more.

True cloud computing services are designed to maximize uptime and possibly even eliminate downtime. That is, their goal is to keep everything running for everyone who uses them, even if some component in the cloud stops working for some reason. The failure of any one server or any one piece of networking equipment or even an earthquake or a total power failure at one data center should not interrupt service to the users because the same data and applications and operations are available on a number of servers that are globally scattered. The buzzword in cloud computing is **REDUNDANCY**: multiple servers in multiple locations, all configured in such a way that the failure of any one device or even an entire location will not interrupt service to the users. Instead, the workload is simply redistributed to other servers.

Let me use one popular service as an example. **Gmail** is a service that runs on Google's many

servers in the cloud. We do not need to install an email program in our own computers in order to use Gmail; the required software actually resides in the Gmail servers. These servers are located in multiple data centers around the world. When you go to check your in-box in Gmail, you simply connect to *http://www.gmail.com.* You don't know where that mail server is and you don't care.

Let's say that you get connected to a Gmail server in California. Assuming that all goes well, you will probably remain connected to that one server for the duration of your session. However, if that server encounters a hardware error or even powers off in the middle of reading your in-box, you are automatically transferred to a different server (probably in the same data center) that has a duplicate copy of your information. You won't notice anything unusual.

If the entire data center in California experiences a power outage (or an earthquake), the entire data center will be knocked offline. However, your in-box session will automatically be reassigned to a different server in a different location, possibly in Singapore. Again, you won't notice any change. Everything appears to function normally on your computer's screen.

This is cloud computing.

When first introduced in 1993, the World Wide Web was ONE-WAY, similar to a newspaper or an encyclopedia. That is, we could find information and display it on our computer screens or print it on our local printers. Information was sent from web servers to the individuals.

As the years went by and clever programmers created new applications, the Web became a two-way medium for sending and retrieving data. This was the beginning of "the cloud."

For some number of years, the Web was primarily a means of sending data back and forth. If we wanted to run a program, such as a word processor or a spreadsheet or a data-base program, we had to install the required software on our local computers or on a server owned by our employers or some other large organization. In the case of a corporation, one server might serve a handful of users or perhaps a few hundred people. There were only few large databases that served perhaps a thousand simultaneous users. If those servers connected to the Web, a user outside of the corporation might be able to see the results of the corporate computing. For example, we could see items on eBay or flight schedules on Expedia, make a purchase, and get confirmation of that purchase. Similarly, we could search for family records, receive data back, and transcribe that data into a genealogy program on our home computers. The Web carried the information back and forth, but the programming was done on one end of the transmission or the other, traceable to a specific location.

The third phase of implementing the cloud was when programmers developed methods of running all these programs and more on distant servers. As the technology evolved, it was no longer necessary to install programs on each user's computer. Instead, multi-user programs were installed on distant servers, and the users accessed those programs via the Internet. Today we can use small, low-powered computers to access and run powerful programs that are installed on distant servers, which makes computing more powerful and less expensive for all of us.

As an example, the new Chromebook laptop computers have almost no programs installed on their tiny hard drives. The same is true for tablet computers and smartphones, which don't even contain hard drives. All they require is an Internet connection (usually wireless), and they can run all sorts of word processors, spreadsheets, computer games, publishing software, photo editing programs, and more. Yes, they can also run genealogy programs to store the results of your family tree research. The programs are installed on the distant server, not on the local computer.

Another example is **Waze**, a very popular mapping and GPS application used when traveling by automobile. With Waze, we do not need to install huge databases in our cell phones, databases that contain information about every street, highway, intersection, fast food restaurant, hospital, gas station, traffic jam, and more in the world. We also do not need to install sophisticated navigation software to use all that data as all data processing is performed by systems in the cloud. We also do

not need to have powerful (and expensive) central processors in our cell phones in order to calculate the best route to travel. All that is data processing is again performed by systems in the cloud.

This is cloud computing.

I use a four-ounce cell phone that is with me wherever I travel. By connecting to the Internet and accessing programs and databases in the cloud, I have computing power that would have cost thousands of dollars only twenty years ago. The same is true of the tablet computers and Chromebooks. These modern marvels are even battery-powered and connect to the cloud via wireless connections.

In effect, we are using low-powered, low-cost devices that provide "windows" into very fast, expensive computers located in data centers around the world. We are taking advantage of an elaborate collection of hardware, software, data, and networks that work together, invisible to us. These net-worked collections of systems also provide super high reliability, frequently displaying no downtime to the users for a year or more. It is this collection that we have come to know as "the cloud."

One common question is, "How do I connect to the cloud?" If you have an Internet connection, you already have access to the cloud. In turn, the cloud gives you access to thousands of services, and all of them are available through the Internet.

The cloud also may be envisioned as the next evolution beyond the World Wide Web. While the original World Wide Web delivered information one-way to the user, the cloud does all that and more. The cloud provides two-way data as well as multi-user and even collaborative applications. Do you use Google Docs? If so, you are already using the cloud. Do you use BillionGraves or Find-A-Grave? If so, you are already using the cloud. Do you pay bills online? Do you access your bank or investment account online? If so, you are already using the cloud. The same is true for Facebook, Gmail, Flickr, Shutterfly, Mozy, Carbonite, and dozens of other cloud-based services.

Cloud computing offers other benefits as well. Not too long ago, many of us worried about losing our documents, photos, and files if something bad happened to our computers, such as a hard drive crash or a virus. Today, our data is migrating beyond the boundaries of our personal computers. Instead, we're moving our data online—into "the cloud"—for security reasons. If you upload your photos, store critical files online, and use a web-based email service like Gmail or Yahoo! Mail, an 18-wheel truck could run over your laptop, and all your data would still remain safely stored on the web, accessible from any Internet-connected computer, anywhere in the world. Not only is your information available online but it is also available on MULTIPLE REDUNDANT computers in the cloud. If a different 18-wheeler crashes through the wall of a data center and runs into one of the servers hosting your data, the information is still saved and secure in the cloud service's other servers in other locations that are sharing the workload.

There is no sharp dividing line between the World Wide Web and "the cloud." In fact, the cloud is an evolution that happened over a period of several years. Some of the early applications are tough to define as to whether they are Web applications or cloud-based applications. However, Gmail,

Facebook, Flickr, and Shutterfly are true cloud-based applications.

Another buzzword you will hear often is "app." In fact, an app is simply shorthand for an application. Applications are also called programs or software. Traditionally, they've been designed to perform computer-intensive tasks, such as accounting or word processing. In the online world of web browsers and smart phones, apps are usually smaller, nimbler programs focused on a single task.

Some cloud-based services require installation of a simple app in the user's cell phone, tablet, laptop, or desktop computer. In other cases, all you need to do is open a web browser and go to a specific web address to access the cloud-based service.

Another advantage of cloud applications is that the software resides on distant servers, not in our local computers. Software upgrades and troubleshooting are performed by computer professionals employed by the cloud companies, not by the individual user. Which version of YouTube am I using today? What about tomorrow? The answer: I am always using the latest version. Web apps get updated on the servers, so everyone always accesses the latest version, with all the newest features and improvements. There is no need for anyone to manually upgrade to a new version. All the work and storage space is provided far from the users and invisible to them.

The cloud-based services usually work on all sorts of computers. In traditional computing, most programs work only on Windows or only on Macintosh computers. In contrast, the web is an open platform. Anyone can reach it from a browser on any web-connected device, regardless of whether it's a desktop computer, laptop, or mobile phone. That means I can use my favorite web apps even if I'm using my friend's laptop or a computer at an Internet cafe. I can even access most everything from a Windows computer at home, from a Macintosh laptop, from an Apple iPad running the iOS operating system, from an Android smartphone, or from a low-cost Chromebook laptop.
NOTE: There are a few exceptions. A very few cloud applications may communicate only with a Windows computer or an Android tablet or an iPhone. However, the overwhelming majority of cloud-based applications are device-agnostic; that is, they run on any modern computing device.

Cloud applications are also more resistant to viruses and other bad things. While nothing is ever perfect and no one can ever guarantee freedom from such problems, cloud apps historically have had fewer problems with viruses and similar malevolent software than have Windows computers. In short, cloud applications are more resistant to attacks than are programs and data stored in a local computer.

Another huge advantage of cloud applications is that the data and the application are stored in multiple large data centers managed by computer professionals. Data is backed up regularly, unlike many home computers. If your laptop is stolen from your automobile, the thief who steals it will not find your personal information on the laptop's hard drive.

Speaking of thieves, your information stored in the cloud is typically more secure than is data stored on your desktop or laptop computer. Locally, visitors to your house such as plumbers, electricians, carpenters, your child's or grandchild's friends, or the shady brother-in-law you never quite trusted, all can easily gain access to the data stored in your computer.

In contrast, most cloud-based services use industrial-grade encryption to keep your information safe from prying eyes. Banks were the first to adopt such security methods, but the same technology is now used by most, although not all, cloud-based services. In most cases, your personal information is safer when stored in the cloud than it is when stored on your own computer's hard drive.

This concludes the introduction to cloud computing. In Part 2 of this article, I will describe a number of cloud-based services and applications designed for genealogical uses.
To see a list of some of the available cloud services, look at *https://www.intelcloudfinder.com/cloud-providers.*